

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/51512/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

de Miranda, Noel F. C. C., Peng, Roujun, Georgiou, Konstantinos, Wu, Chenglin, Sorqvist, Elin Falk, Berglund, Mattias, Chen, Longyun, Gao, Zhibo, Lagerstedt, Kristina, Lisboa, Susana, Roos, Fredrik, van Wezel, Tom, Teixeira, Manuel R., Rosenquist, Richard, Sundstrom, Christer, Enblad, Gunilla, Nilsson, Mats, Zeng, Yixin, Kipling, David Glyn and Pan-Hammarstrom, Qiang 2013. DNA repair genes are selectively mutated in diffuse large B cell lymphomas. *Journal of Experimental Medicine* 210 (9) , pp. 1729-1742. 10.1084/jem.20122842 file

Publishers page: <http://dx.doi.org/10.1084/jem.20122842>
<<http://dx.doi.org/10.1084/jem.20122842>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



DNA repair genes are selectively mutated in diffuse large B cell lymphomas

Noel FCC de Miranda,¹ Roujun Peng,^{1,2} Konstantinos Georgiou,¹ Chenglin Wu,¹ Elin Falk Sörqvist,³ Mattias Berglund,⁴ Longyun Chen,⁵ Zhibo Gao,⁵ Kristina Lagerstedt,⁶ Susana Lisboa,⁷ Fredrik Roos,³ Tom van Wezel,⁸ Manuel R. Teixeira,⁷ Richard Rosenquist,³ Christer Sundström,³ Gunilla Enblad,⁴ Mats Nilsson,^{3,9} Yixin Zeng,² David Kipling,¹⁰ and Qiang Pan-Hammarström¹

¹Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet at Karolinska University Hospital, 14186 Huddinge, Sweden

²State Key Laboratory of Oncology in South China and Department of Medical Oncology, Sun Yat-Sen University Cancer Center, Guangzhou 510275, China

³Department of Immunology, Genetics and Pathology and ⁴Department of Radiology, Oncology and Radiation Science, Uppsala University, 75185 Uppsala, Sweden

⁵BGI-Shenzhen, Shenzhen 518083, China

⁶Department of Molecular Medicine and Surgery, Karolinska University Hospital, 17176 Solna, Sweden

⁷Department of Genetics, Portuguese Oncology Institute, Porto, and Abel Salazar Biomedical Sciences Institute (ICBAS), Porto University, 4200-072 Porto, Portugal

⁸Department of Pathology, Leiden University Medical Center, 2300RC Leiden, Netherlands

⁹Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden

¹⁰Institute of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4YU, Wales, UK

DNA repair mechanisms are fundamental for B cell development, which relies on the somatic diversification of the immunoglobulin genes by V(D)J recombination, somatic hypermutation, and class switch recombination. Their failure is postulated to promote genomic instability and malignant transformation in B cells. By performing targeted sequencing of 73 key DNA repair genes in 29 B cell lymphoma samples, somatic and germline mutations were identified in various DNA repair pathways, mainly in diffuse large B cell lymphomas (DLBCLs). Mutations in mismatch repair genes (*EXO1*, *MSH2*, and *MSH6*) were associated with microsatellite instability, increased number of somatic insertions/deletions, and altered mutation signatures in tumors. Somatic mutations in nonhomologous end-joining (NHEJ) genes (*DCLRE1C/ARTEMIS*, *PRKDC/DNA-PKcs*, *XRCC5/KU80*, and *XRCC6/KU70*) were identified in four DLBCL tumors and cytogenetic analyses revealed that translocations involving the *immunoglobulin-heavy chain* locus occurred exclusively in NHEJ-mutated samples. The novel mutation targets, *CHEK2* and *PARP1*, were further screened in expanded DLBCL cohorts, and somatic as well as novel and rare germline mutations were identified in 8 and 5% of analyzed tumors, respectively. By correlating defects in a subset of DNA damage response and repair genes with genomic instability events in tumors, we propose that these genes play a role in DLBCL lymphomagenesis.

CORRESPONDENCE

Qiang Pan-Hammarström:
Qiang.Pan-Hammarstrom@ki.se

Abbreviations used: AID, activation-induced cytidine deaminase; BER, base-excision repair; BL, Burkitt lymphoma; CSR, class switch recombination; DDR, DNA damage response; DLBCL, diffuse large B cell lymphoma; DSB, double-strand break; FA, Fanconi anemia; FISH, fluorescent in situ hybridization; FL, follicular lymphoma; HR, homologous recombination; MAF, minor allele frequency; MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stable; NER, nucleotide excision repair; NHEJ, non-homologous end joining; ROI, regions of interest; SHM, somatic hypermutation.

Normal lymphocyte development and function relies on the successful rearrangement and modification of antigen receptor genes. Diversity of the B cell receptor is largely provided by V(D)J recombination where the variable (V), diversity (D), and joining (J) segments of the *immunoglobulin (IG)* loci are joined in a combinatorial manner (Jung et al., 2006). After antigen

experience, the *IG* loci are further modified by somatic hypermutation (SHM) and class switch recombination (CSR). The recombination-activating proteins 1 and 2 (RAG1 and RAG2) introduce double-strand breaks (DSBs) at recombination signal sequences located around

© 2013 de Miranda et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

Noel FCC de Miranda and Roujun Peng contributed equally to this paper.

Supplemental Material can be found at:
<http://jem.rupress.org/content/suppl/2013/08/16/jem.20122842.DC1.html>

the V, D, and J genes during V(D)J recombination, whereas activation-induced cytidine deaminase (AID) initiates SHM and CSR by deaminating cytosines to uracils at the V and switch (S) regions of the *IG* loci (Jung et al., 2006; Di Noia and Neuberger, 2007).

A myriad of DNA damage response (DDR) and repair proteins mediate and regulate *IG* diversification processes. AID activity provokes guanine/uracil mismatches that are processed by proteins of the base-excision repair (BER) pathway (UNG, APEX1) and mismatch repair (MMR) pathway (MSH2/MSH6, MLH1/PMS2, and EXO1; Di Noia and Neuberger, 2007; Stavnezer et al., 2010). While in the context of CSR such mismatches lead to the generation of DNA DSBs, during SHM, AID activity preferentially results in the establishment of point mutations, although small duplications and deletions may also occur. The resolution of DSBs in V(D)J recombination and CSR is primarily mediated by the non-homologous end-joining (NHEJ) pathway that becomes activated by DDR proteins such as ATM and Nibrin (NBN; Kotnis et al., 2009). The x-ray repair cross-complementing proteins 4 (XRCC4), XRCC5 (Ku80), and XRCC6 (Ku70) and DNA ligase 4, Artemis, DNA-PKcs, and Cernunnos (XLF) proteins are considered to be the core members of NHEJ (Lieber, 2010).

Most B cell neoplasms are thought to originate from antigen-experienced B cells, as tumor cells display SHM at the *IG* V genes (Klein and Dalla-Favera, 2008). Furthermore, a role for *IG* diversification mechanisms in the propagation of genomic instability in mature B cell lymphomas is supported by numerous observations. Non-*IG* genes, including proto-oncogenes such as *MYC*, *BCL6*, *PIM1*, *RHOH*, or *PAX5* are often targeted by SHM, especially in diffuse large B cell lymphomas (DLBCLs; Pasqualucci et al., 2001), one of the most common and aggressive mature B cell lymphoma subtypes. Chromosomal translocations involving the *IG* loci, with breakpoints in S regions and SHM targets, are also a hallmark of mature B cell lymphomas (Küppers and Dalla-Favera, 2001; Lenz et al., 2007). Furthermore, AID has been shown to be essential for the occurrence of *c-myc/IgH* translocations and oncogene-driven induction of germinal center-derived lymphomas in mice (Ramiro et al., 2004; Pasqualucci et al., 2008). Of note, the t(14;18) translocation, involving the *IGH* and the *BCL2* loci, characteristic of follicular lymphomas (FLs), is considered to be derived from defective V(D)J recombination processes (Küppers and Dalla-Favera, 2001).

Despite the crucial role of DDR and repair proteins during antibody diversification processes, there is lack of evidence that supports their direct involvement in the propagation of genomic instability in human B cell lymphomas. In contrast, individuals with biallelic germline mutations in some of the DNA repair genes that encode proteins involved in *IG* diversification processes often display an increased risk for development of lymphoid malignancies in addition to immunodeficiency (de Miranda et al., 2011). In this study, we systematically analyzed the coding regions of key DDR and repair genes that have been associated, or could potentially be

associated, with *IG* gene diversification processes in a set of mature B cell lymphomas, with a focus on DLBCL. The defects in a subset of DDR and repair genes identified here, and their association with genomic instability phenotypes, support their role in the tumorigenesis of DLBCL.

RESULTS

Performance of target enrichment and SOLiD sequencing

73 DDR and repair genes (from this point forward referred to as DNA repair genes), belonging to the DDR, BER, nucleotide excision repair (NER), MMR, NHEJ, Fanconi anemia (FA), and homologous recombination (HR) pathways (Table S1), were selectively sequenced in 29 mature B cell lymphomas, including 22 DLBCLs, 5 FLs, 2 Burkitt lymphomas (BLs), and their respective paired blood samples. As a positive control, the *TNFAIP3* gene, encoding a known tumor suppressor (A20) in DLBCL (Compagno et al., 2009), was also analyzed in these samples. Target enrichment was accomplished by the Selector technology and samples were sequenced with the SOLiD platform. More than 10^6 SOLiD short-reads were generated per sample with >90% of reads being specific for the defined regions of interest (ROI; Table S2). Mean sequencing depth was >200×, with >84 and 90% of the ROI being covered by at least 10 and 5 SOLiD sequencing reads, respectively (Table S2). Validation of the sequencing procedure and variant calling methods were performed by processing a HapMap sample (NA19240) in parallel with the clinical samples. A concordance of 99.6% was achieved in the 499 heterozygous positions covered by our Selector design in the HapMap sample.

Detection of somatic and germline mutations in DNA repair genes

To detect germline as well as somatic mutations, sequences from both normal and tumor samples were compared with the human genome reference (hg18), and subsequently, minor allele frequencies (MAFs) at heterozygous positions were compared between tumor and normal pairs. Filtering of known polymorphisms was performed against dbSNP132 and variant novelty verified with dbSNP137. By combining both analysis and removing overlapping results, 442 single nucleotide variants (SNVs) were identified across all samples. Approximately half of these variants passed visual inspection, which allowed the elimination of false positive results derived from library preparation, sequencing, and/or alignment artifacts (Fig. S1). Such artifacts were normally located toward the end of the sequencing reads or were variants supported by identical sequencing reads. 124 heterozygous SNVs resulted in nonsynonymous amino acid changes that included novel germline and somatic mutations as well as rare germline variants with an MAF below 0.01 (Fig. S1). All novel germline and somatic mutations and rare variants were validated by Sanger sequencing and are presented in Table S3. Detection of indels proved to be difficult, and an alternative alignment of SOLiD short-reads was produced with BFAST that allowed the identification of four somatic deletions in four different

Table 1. Somatic mutations discovered by targeted SOLiD sequencing

Gene	Base change	AA change	Percentage of sequencing reads	SIFT	Polyphen
			%		
<i>ATM</i>	c.3616T>G	p.L1206V	10	Tolerated	Benign
<i>ATM</i>	c.3646T>C	p.Y1216H	12	Tolerated	Damaging (probably)
<i>BRCA2</i>	c.2546T>C	p.V849A	33	Tolerated	Benign
<i>BRCA2</i>	c.9966G>T	p.M3322I	36	Tolerated	Damaging (probably)
<i>CHEK2</i>	c.1007A>T	p.Q336L	54	Damaging	Benign
<i>DCLRE1C</i>	c.1628_1632delTAACA	p.(I543Rfs*12)	10	-	-
<i>DDB1</i>	c.26C>G	p.A9G	86	Damaging	Damaging (possibly)
<i>EXO1</i>	c.1412A>G	p.K471R	18	Tolerated	Benign
<i>MSH2</i>	c.1618A>G	p.S540G	20	Tolerated	Benign
<i>MSH2</i>	c.2497C>G	p.L833V	50	Damaging	Benign
<i>MSH6</i>	c.1628A>G	p.K543R	29	Damaging	Benign
<i>PARP1</i>	c.1910_1922delinsGTT	p.(K637Sfs*13)	18	-	-
<i>PRKDC</i>	c.10252T>C	p.F3418L	31	Damaging	Damaging (probably)
<i>TNFAIP3</i>	c.1877_1878delTG	p.(C627Ffs*44)	20	-	-
<i>TP53</i>	c.400T>G	p.F134V	47	Damaging	Damaging (probably)
<i>TP53</i>	c.817C>T	p.R273C	67	Damaging	Damaging (probably)
<i>TP53</i>	c.818G>A	p.R273H	38	Damaging	Damaging (probably)
<i>XRCC5</i>	c.847A>G	p.T283A	29	Damaging	Damaging (probably)
<i>XRCC6</i>	c.800delT	p.(I267Kfs*2)	28	-	-

genes across three independent tumors. All nonsynonymous somatic mutations were detected exclusively in DLBCL cases (Table S3). No somatic mutations were found in FL and BL samples despite comparable sequencing performances.

DNA repair pathways targeted by mutations in DLBCL

19 somatic mutations were discovered by SOLiD sequencing, distributed in 10 DLBCL tumors (Table 1 and Table S3). These were more frequent in DDR factors, including the classical tumor suppressor genes *ATM* and *TP53*, and the *CHEK2* and *PARP1* genes that are novel mutation targets in DLBCL. Recurrent alterations in MMR genes (*EXO1*, *MSH2*, and *MSH6*) and members of the NHEJ pathway (*DCLRE1C/ARTEMIS*, *PRKDC/DNA-PKcs*, *XRCC5/KU80*, and *XRCC6/KU70*) were also observed (Fig. 1 and Table 1). Additionally, two somatic mutations were found in *BRCA2* (HR pathway) and one mutation in the *DDB1* (NER pathway) and *TNFAIP3* genes (Table 1). To the best of our knowledge, and as observed for *CHEK2* and *PARP1*, nonsynonymous somatic mutations in *DCLRE1C*, *EXO1*, and *XRCC5* have not been previously reported in DLBCL. The somatic mutations discovered in the *DCLRE1C*, *PARP1*, *XRCC6*, and *TNFAIP3* genes were frameshift deletions, whereas the remaining somatic mutations led to amino acid substitutions (Table 1). The SIFT and Polyphen-2 prediction tools provided a discordant functional prediction for five missense mutations (Table 1), which may be due to the use of different algorithms and sequence alignments (Hicks et al., 2011).

Based on the percentage of sequencing reads, the majority of somatic mutations were considered to be present in dominant tumor cell clones, except for the ones targeting *ATM*

and *EXO1* which were represented by <20% of sequencing reads (Table 1). The low percentage of sequencing reads that supported the deletions at the *DCLRE1C* and *PARP1* genes probably resulted from the problematic alignment of SOLiD reads with deletions of five or more nucleotides. Nevertheless, these deletions were readily validated by Sanger sequencing and are likely to be present in dominant tumor cell clones in a heterozygous state.

The mean frequency of nonsynonymous, somatic mutations in DNA repair genes was 4.16 mutations/Mb of target sequence (protein coding). To compare the mutation frequencies in DNA repair genes with the ones found in the entire coding genome (exome), we performed exome sequencing in 17 out of the 22 DLBCL cases that underwent Selector/SOLiD sequencing. Although biases could arise from the use of different techniques, the somatic mutations in DNA repair genes discovered by exome sequencing largely overlapped with the ones detected by Selector/SOLiD and translated into a mutation frequency of 4.21 mutations/Mb. This frequency was similar to the ones determined for the entire coding genome (3.15 mutations/Mb, 20,930 genes) and for specific groups of genes such as kinases (3.71 mutations/Mb, 507 genes) or transcription factor genes (3.44 mutations/Mb, 1,645 genes). Of note, the latter gene groups encompass frequent mutation targets in DLBCL such as *PIM1* (kinase) or *BCL6* and *TP53* (transcription factors). After excluding mutations in the classical tumor suppressor *TP53*, the somatic mutation frequency in DNA repair genes detected by exome was reduced (3.01 mutations/Mb) but remained comparable to the ones derived from the entire coding genome and other gene groups.

Table 2. Summary of variations discovered in *CHEK2* in an expanded DLBCL cohort ($n = 235$)

DNA	Protein	Novel/SNP	Mutated tumors	Somatic/germline	Ethnic origin	MAF controls†	MAF 1000 genomes§	SIFT	Polyphen
c.319+2T>A	-	Novel	1/171	ND	Swedish	0	0	-	-
c.876dupT	p.D293X	Novel	1/171	Somatic	Swedish	0	0	-	-
c.1007A>T	p.Q336L	Novel	1/64	Somatic	Chinese	0	0	Damaging	Damaging (possibly)
c.1091T>C	p.I364T	Novel	1/171	ND	Swedish	0	0	Damaging	Damaging (probably)
c.7C>T	p.R3W	rs199708878	1/171	Germline	Swedish	0	0	Damaging	Damaging (probably)
c.349A>G	p.R117G	rs28909982	1/171	ND	Swedish	0	0	Damaging	Damaging (probably)
c.444+1G>A	-	Reported	1/171	Germline	Swedish	0	0	-	-
c.1312G>T	p.D438Y	rs200050883	2/164	ND	Swedish	0	0	Damaging	Damaging (probably)
c.1582G>A	p.E528K	rs138040612	1/64	Germline	Chinese	0	0	Tolerated	Damaging (possibly)
c.470T>C	p.I157T	rs17879961	3/171	Germline (1)	Swedish	0.008	0.007	Tolerated	Damaging (possibly)
c.538C>T	p.R180C	rs77130927	1/171	ND	Swedish	0.006	0.001	Damaging	Damaging (possibly)
c.1100delC	p.T367Mfs*15	Reported	3/163	ND	Swedish	0.006	0	-	-
c.1111C>T	p.H371Y	Reported	2/64	Somatic/germline	Chinese	0.004	0	Damaging	Benign

The expanded DLBCL cohort was comprised of 171 Swedish and 64 Chinese cases. Population-specific MAFs for MAF controls were calculated by typing 1,030 Swedish or 429 Chinese controls, with an exception for p.D438Y, p.E528K, and c.1100delC variants, which were screened by Sanger in 703, 361, and 345 controls, respectively. For MAF, data from 1,000 genomes were available for 379 European and 197 Chinese controls. ND, not determined. (1), germline origin was confirmed in one patient.

19 mature B cell lymphoma patients carried novel non-synonymous germline mutations in different DNA repair genes (Table S3). These were more frequent in members of the MMR pathway (Fig. 1). Three mutations were identified in the *MSH3* gene, whereas two mutations were detected in the *MLH3* gene in four DLBCL patients (Table S3). Furthermore, nonsynonymous germline mutations in *MSH6*, *PMS1*, and *PMS2* were also identified in DLBCL patients, whereas one FL patient carried a novel mutation in *MSH2*. Although somatic mutations

were rare in HR and FA proteins, several germline mutations were identified in these pathways. Alterations in genes functionally related to the DDR pathway were also common. One DLBCL and two FL patients carried novel germline mutations in the *MDC1* gene that encodes a core DDR protein. 21 rare variants (with MAF < 0.01) were identified in 17 mature B cell lymphoma patients. Of note, four patients (three DLBCL and one FL) carried rare nonsynonymous substitutions in the gene encoding TP53BP1, another essential DDR protein (Table S3).

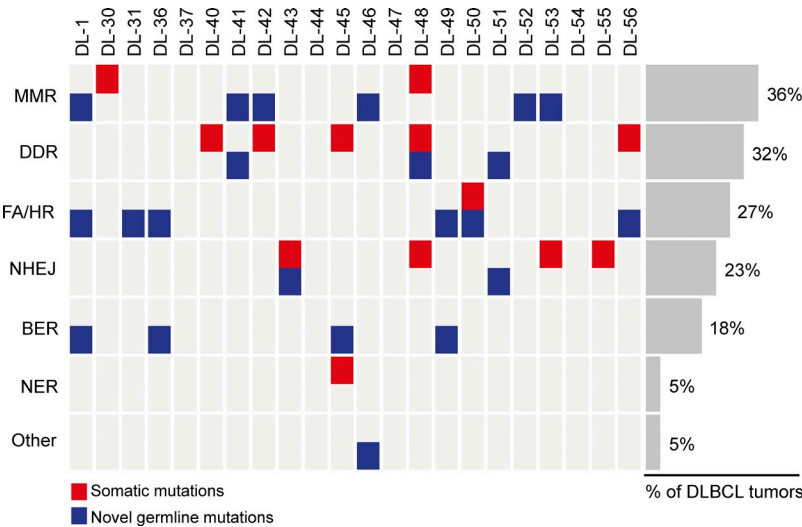


Figure 1. DNA repair pathways affected by somatic and novel germline mutations in DLBCL. Somatic and novel germline mutations together were most frequent in MMR genes, followed by general DDR genes and NHEJ genes. To a lower extent, mutations were detected in FA, HR, BER, and NER pathways. The "Other" category refers to the p.R335W mutation in the *DNIT* gene, which encodes for a protein involved in the addition of nucleotides at junctions of rearranged *IGH* and *TCR* gene segments during V(D)J recombination. The percentage of DLBCL tumors affected by somatic or novel germline mutations in different DNA repair pathways is displayed. Mutations were discovered by SOLiD and validated by Sanger sequencing.

Table 3. Summary of variations discovered in *PARP1*, *MSH2*, *DDB1*, *DCLRE1C*, *MLH3*, and *XRCC5* after sequencing of expanded DLBCL cohorts

Gene	DNA	Protein	Method	Novel/SNP	Somatic/ germline	Ethnic origin	MAF 1,000 genomes†	SIFT	Polyphen
<i>PARP1</i>	c.1897A>T	p.K633X	Sanger	Novel	Somatic	Chinese	0	-	-
	c.1910_1922delinsGTT	p.(K637Sfs*13)	Selector & Sanger	Novel	Somatic	Chinese	0	-	-
	c.2326A>G	p.S776G	Selector & Sanger	Novel	Germline	Chinese	0	Damaging	Damaging (possibly)
	c.2756T>C	p.I919T	Sanger	Novel	ND	Chinese	0	Damaging	Damaging (possibly)
<i>MSH2</i>	c.2789A>G	p.Y930C	Sanger	rs189410139	ND	Chinese	0.003	Damaging	Damaging (probably)
	c.3037C>A	p.L1013M	Sanger	rs138906127	ND	Swedish	0.001	Tolerated	Damaging (possibly)
	c.1316C>T	p.P439L	Sanger	Novel	ND	Chinese	0	Damaging	Damaging (probably)
	c.1619A>G	p.S540G	Selector & Sanger	Novel	Somatic	Chinese	0	Tolerated	Benign
	c.2498C>G	p.L833V	Selector & Sanger	Novel	Somatic	Chinese	0	Damaging	Benign
<i>DDB1</i>	c.505A>G	p.I169V	Sanger	rs63750716	ND	Chinese	0	Tolerated	Benign
	c.26C>G	p.A9G	Selector	Novel	Somatic	Chinese	0	Damaging	Damaging (probably)
	c.2375_2376dupTT	p.(I793Lfs*75)	454 amplicon	Novel	ND	Chinese	0	-	-
<i>DCLRE1C</i>	c.3360G>A	p.M1120I	454 amplicon	Novel	ND	Swedish	0	Tolerated	Benign
	c.536G>A	p.R179Q	454 amplicon	rs150854849	ND	Swedish	0	Damaging	Damaging (probably)
<i>MLH3</i>	c.1628_1632delTAACA	p.(I543Rfs*12)	Selector	Novel	Somatic	Chinese	0	-	-
	c.119G>A	p.C40Y	Selector	Novel	Germline	Chinese	0	Damaging	Damaging (probably)
	c.365A>G	p.K122R	454 amplicon	Novel	ND	Chinese	0	Damaging	Damaging (probably)
	c.2221G>T	p.V741I	454 amplicon	rs28756990	ND	Swedish	0.004	Tolerated	Damaging (possibly)
	c.2837C>T	p.S946F	Selector	rs201441389	Germline	Chinese	0.003	Tolerated	Damaging (possibly)
<i>XRCC5</i>	c.2964C>G	p.I988M	Selector	Novel	Germline	Chinese	0	Damaging	Benign
	c.3331C>T	p.L1111F	Selector	Novel	Germline	Chinese	0	Damaging	Damaging (probably)
	c.847A>G	p.T283A	Selector	Novel	Somatic	Chinese	0	Damaging	Damaging (probably)
	c.1046C>A	p.S349Y	454 amplicon	Novel	Germline	Chinese	0	Damaging	Damaging (probably)
	c.1522A>G	p.I508V	454 amplicon	rs2287558	ND	Swedish	0.003	Tolerated	Benign

For expanded DLBCL cohort, *PARP1* and *MSH2* genes were analyzed in 130 (95 Chinese and 35 Swedish) and 94 (59 Chinese and 35 Swedish) DLBCL samples, respectively, whereas *DDB1*, *DCLRE1C*, *MLH3*, and *XRCC5* genes were sequenced in 48 (25 Chinese and 23 Swedish) DLBCL samples. MAF 1,000 genomes data were available for 379 European and 197 Chinese controls. ND, not determined.

Screening of mutation targets in expanded cohorts of DLBCL tumors

Six mutated genes discovered by SOLiD sequencing were selected for further investigation in an extended cohort of DLBCL tumors. These include *CHEK2*, *PARP1*, *DDB1*, and *XRCC5*, which were novel mutation targets for DLBCL, and *MSH2* and *MLH3*, which were recurrently affected by somatic and/or germline mutations, respectively.

The *CHEK2* gene was Sanger sequenced in a total of 235 DLBCL samples, including the 22 DLBCL samples that underwent targeted SOLiD sequencing. In addition to the p.Q336L somatic mutation and the p.E528K germline variant, identified by SOLiD, 11 novel mutations and previously reported rare variants were detected, distributed in 17 tumors that did not undergo SOLiD sequencing (Table 2). The novel mutations identified in *CHEK2* included a somatic frameshift insertion (p.D293X), a splice-site mutation (c.319+2T>A), and a missense mutation (p.I364T) located in the kinase domain of *CHEK2* (Table 2). Five rare variants,

including p.E528K, were specific for the patient cohort, as they were not identified in large groups of controls (Table 2). Of these, the c.444+1G>A splice site mutation has been previously reported to lead to aberrant splicing of *CHEK2* transcripts, whereas the p.R117G and p.D438Y substitutions were shown to have deleterious effects on the function of *CHEK2* (Sodha et al., 2006; Bell et al., 2007). Four other rare variants were found in both tumor and control samples, including a c.1100delC deletion and a p.I157T substitution (Table 2) that was previously shown to have a damaging effect on protein function (Falck et al., 2001). Sequencing of the entire coding regions of the *CHEK2* gene was also performed in 90 samples derived from healthy blood donors, where the p.R180C and p.I157T substitutions, previously identified by genotyping, were the only variants found in one and two samples, respectively. Thus, the discovery of novel mutations or certain rare variations was restricted to the DLBCL patient group, further suggesting their association with the disease.

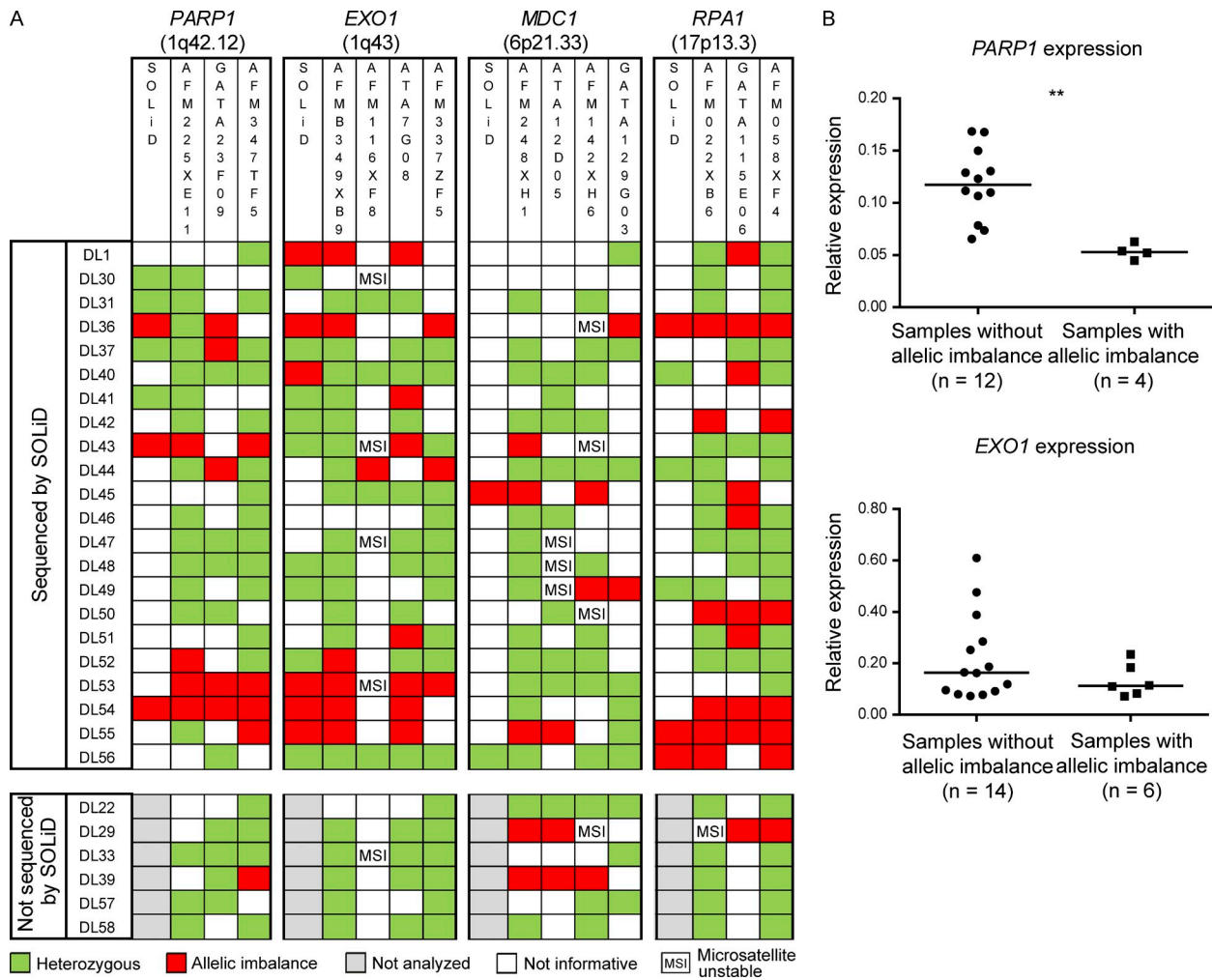


Figure 2. Allelic imbalance at *PARP1*, *EXO1*, *MDC1*, and *RPA1* loci in DLBCL samples and association with *PARP1* expression. (A) Allelic imbalance events detected by SOLiD were validated by the use of polymorphic markers and fragment analysis in 22 DLBCL samples. Six additional samples that did not undergo SOLiD sequencing were also investigated. Some markers were not informative, because of either germline homozygosity or MSI in the tumor samples. Markers displaying allelic imbalance were analyzed in two independent experiments. (B) Expression of *PARP1* and *EXO1* was determined by quantitative real-time PCR in samples with and without allelic imbalance at the respective loci. Medians are depicted and the statistical difference between groups was assessed with the Mann-Whitney test (**, $P < 0.01$). Two independent experiments were performed and a representative experiment is shown.

Sanger sequencing of the *PARP1* gene in 130 DLBCL samples, including the tumors that underwent SOLiD sequencing, allowed the discovery of two additional novel nonsynonymous mutations: a STOP-gain mutation located upstream of the catalytic domain of the PARP1 protein (p.K633X) and a p.I919T substitution located in the catalytic domain of PARP1 (Table 3). Two rare variants (p.Y930C and p.L1013M) were also identified in two different tumors. Overall, variations in the *PARP1* gene were identified in 5% of samples analyzed.

The *MSH2* gene was also Sanger sequenced in 94 DLBCL tumors, including the samples that underwent SOLiD sequencing. In addition to the two somatic mutations (p.S540G and p.L833V) discovered by the latter approach, another two nonsynonymous variants were found in two tumor samples:

a novel missense mutation (p.P439L) and a rare variant (p.I169V; Table 3).

The screening of *DCLRE1C*, *DDB1*, *MLH3*, and *XRCC5* genes in 48 DLBCL samples that did not undergo SOLiD sequencing was performed in a high-throughput manner by combining all PCR reactions in a microfluidic PCR array followed by 454 GS FLX Titanium sequencing. Two novel mutations were identified in the *DDB1* gene, in two different tumors, including a frameshift, truncating mutation in *DDB1* (Table 3). The *MLH3* and *XRCC5* genes also presented one novel mutation each in two additional samples (Table 3). Finally, rare missense variants were identified in *DCLRE1C*, *MLH3*, and *XRCC5* in three tumors, respectively. The *MLH3* missense rare variant (p.V741I) was previously shown to segregate with disease in an endometrial cancer family (Liu et al., 2003).

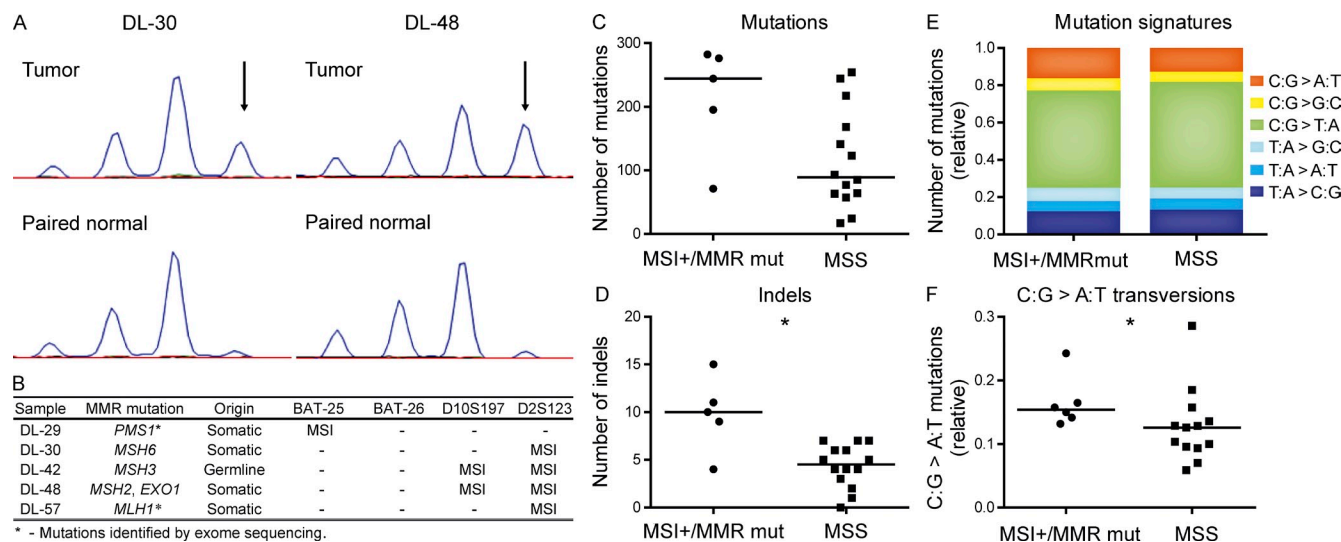


Figure 3. MSI and mutation profiles in MMR-mutated DLBCL. (A) Fragment analysis was performed with microsatellite markers in 28 DLBCL tumors. (B) MSI detection in five DLBCL cases with mutations in MMR genes. MSI detection was confirmed by one independent experiment. (C) The median of the total number of somatic mutations in the coding genome (exome) was higher in MMR-mutated DLBCL tumors that displayed MSI than in MSS tumors. (D) This difference was significant when looking specifically at the number of somatic indels detected by exome sequencing. (E and F) C:G→A:T transversions were significantly enriched in MMR-mutated DLBCL samples displaying MSI. The statistical difference between groups was assessed with the Mann-Whitney test (*, $P < 0.05$).

Detection of allelic imbalance at DNA repair genes loci

Chromosomal alterations could be detected from the data generated by the targeted deep sequencing of DNA repair genes. For that purpose we compared allelic ratios at germline heterozygous SNVs, between tumor and normal samples. The *EXO1* locus was found to be affected by allelic imbalance in six DLBCL tumors, whereas allelic imbalance at the *PARP1* and *RPA1* loci was detected in three DLBCL samples. The *TP53BP1*, *ERCC6*, and *ATR* genes were affected in two samples each and other loci were affected in single DLBCL samples, including the *MDC1* locus.

The allelic imbalance results obtained from SOLiD sequencing were validated by making use of polymorphic markers located around the *EXO1*, *PARP1*, *MDC1*, and *RPA1* loci, and the analysis was extended to a total of 28 DLBCL samples. The majority of results obtained from SOLiD could be confirmed by the employment of polymorphic markers except for one case: *EXO1* in DL-40 (Fig. 2 A). Furthermore, this approach allowed the identification of additional samples affected by allelic imbalance that were not detected by SOLiD data analysis (Fig. 2 A). This can be explained by the fact that the detection of allelic imbalance by the latter required the presence of at least two germline heterozygous SNVs at different loci, which was not always observed. In total, allelic imbalance at *RPA1*, *EXO1*, *MDC1*, and *PARP1* was detected in 25, 21, 18, and 14% of samples, respectively. We subsequently assessed the expression of *RPA1*, *EXO1*, *MDC1*, and *PARP1* and found that the expression of the latter was significantly lower in tumors presenting allelic imbalance (Fig. 2 B). This association might reflect the loss of *PARP1* copy number in samples with allelic imbalance. No significant

differences in terms of gene expression were observed for the other loci, indicating that allelic imbalance did not correlate with gene copy numbers or that the loss of one allele could be compensated at the transcriptional level (Guidi et al., 2004).

Microsatellite instability (MSI) and mutation profiles in MMR-mutated DLBCL

To investigate whether defects in MMR genes translated into a particular form of genetic instability, MSI analysis was performed in 28 DLBCL cases for which paired normal samples were available, including the 22 DLBCL cases that underwent targeted sequencing. Two mononucleotide (BAT-25 and BAT-26) and two dinucleotide (D10S197 and D2S123) microsatellite markers were used (Fig. 3 A). Instability in one or two markers was detected in five DLBCL samples and was almost exclusively restricted to dinucleotide markers. Notably, all samples that displayed instability of microsatellites possessed at least one alteration in an MMR gene (Fig. 3 B). They included the tumor samples DL-30 and DL-48, where somatic mutations in *MSH6*, *MSH2*, and *EXO1* genes were discovered by SOLiD sequencing, and one sample (DL-42) derived from a DLBCL patient that carried a novel germline mutation in *MSH3*. Two additional samples (DL-29 and DL-57) that displayed instability at microsatellites markers did not undergo the current sequencing approach, but have been analyzed by whole-exome sequencing—presented somatic mutations in *PMS1* (p.R847K) and *MLH1* (p.A223D), respectively. The DLBCL sample with a p.P439L mutation in *MSH2* was analyzed with Promega’s MSI analysis system due to the lack of paired normal DNA and presented instability of one (NR-21) out of the five mononucleotide markers.

To investigate whether defects in MMR genes were associated to the mutational load or mutation signatures in DLBCL, whole exome sequencing data were analyzed in the five DLBCL cases in which MSI was detected, in fourteen microsatellite stable (MSS) DLBCL tumors, and in the respective paired blood samples. The total number of somatic mutations detected by exome sequencing was higher in MMR-mutated cases displaying instability of microsatellite markers than in MSS tumors, not reaching but rather close to statistical significance (Mann-Whitney $P = 0.05$, Fig. 3 C). Interestingly, the number of indels was significantly higher in cases with MSI than in MSS tumors (Mann-Whitney $P = 0.02$, Fig. 3 D). This observation suggests that MMR defects in these tumors not only translate into MSI in noncoding regions but also in the occurrence of insertions and deletions within protein-coding exons. Notably, three out of the four deletions identified by targeted sequencing of DNA repair genes occurred in MMR-mutated DLBCL samples displaying MSI. We further investigated whether MMR deficiency was associated with different mutation signatures by comparing the type of mutations observed in MMR-mutated cases displaying MSI with the ones in MSS tumors. Although C:G→T:A transitions were dominant in both groups (Fig. 3 E), MSI-positive DLBCL tumors were significantly enriched with C:G→A:T transversions (Fig. 3 F).

Defects in NHEJ genes and chromosomal translocations in DLBCL

Somatic mutations in NHEJ genes were detected in 4 out of the 22 DLBCL cases that underwent targeted sequencing of DNA repair genes. We further investigated whether mutations in NHEJ genes were associated with genomic instability affecting the *IGH* locus. To this end, we performed fluorescent in situ hybridization (FISH) with *IGH* and *BCL6* dual-color

break-apart probes in 13 DLBCL cases for which tissue sections were available, including 3 NHEJ mutants (DL-43, *XRCC5*; DL-48, *DCLRE1C*; DL-53, *XRCC6*). A split signal affecting one of the *IGH* loci was detected in 2 out of the 13 DLBCL cases that were investigated by FISH (Fig. 4 A). This abnormality was found in approximately half of the nuclei analyzed in each tumor. Remarkably, the two cases where the *IGH* locus was rearranged carried somatic mutations in NHEJ genes: DL-43 (*XRCC5* mutant) and DL-48 (*DCLRE1C* mutant). No chromosomal breakage at the *IGH* locus was detected in the 10 DLBCL samples that contained unmutated NHEJ genes. In spite of the small sample size, the occurrence of *IGH* translocations was significantly different between NHEJ mutants and nonmutants as determined by Fisher's exact test ($P = 0.04$). The DL-43 and DL-48 samples also displayed chromosomal breakage at the *BCL6* locus (Fig. 4 B) but no rearrangement of *IGH* together with *BCL2* or rearrangement of the *MYC* locus (Fig. 4, C and D). We therefore concluded that *BCL6* was the most likely translocation partner of *IGH* in these tumors. Disruption of the *BCL6* locus, without concurrent rearrangement of the *IGH* locus, was detected in the microsatellite unstable DLBCL DL-42 that carried a somatic mutation in *TP53* and a novel germline mutation in *MSH3*.

DISCUSSION

By performing targeted screening of 73 DNA repair genes in a cohort of mature B cell lymphomas, we discovered somatic alterations in several novel and/or potentially functional important mutation targets in DLBCL, including *CHEK2*, *PARP1*, *DDB1*, and several NHEJ genes (*DLRE1C*, *PRKDC*, *XRCC5*, and *XRCC6*), as well as MMR genes (*MSH2*, *MSH6*, and *EXO1*). Although the somatic mutation frequency in the 73 DNA repair gene group was not significantly higher than

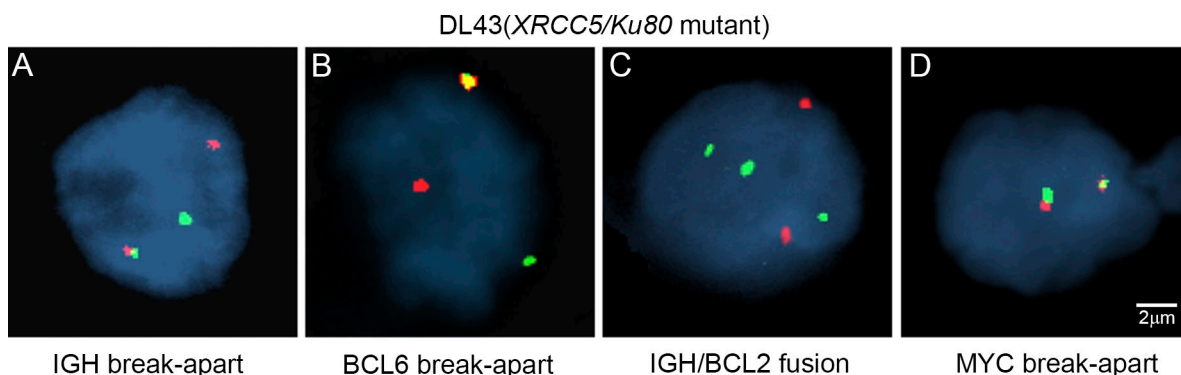


Figure 4. FISH analysis in DLBCL tumors. The integrity of the *IGH* and *BCL6* loci was assessed by FISH in 13 DLBCL tumors that underwent targeted sequencing of DDR and repair genes. Results from DL43 are depicted. (A) Rearrangement of one *IGH* locus was detected in two DLBCL cases (DL43 and DL48), both of which carried a mutation in a NHEJ gene (green, *IGHV* probe; red, *IGH* 3' flanking probe). (B) Rearrangement of one *BCL6* locus was detected in three tumors, including the two cases that displayed disruption of the *IGH* locus (green, *BCL6* 3' flanking probe; red, *BCL6* 5' flanking probe). (C and D) Both cases that displayed concurrent rearrangement of the *IGH* and *BCL6* loci were investigated with an *IGH*/*BCL2* dual fusion translocation probe and a *MYC* break-apart probe to determine the translocation partner of *IGH*. No fusion signal between *IGH* and *BCL2* probes (C) or rearrangement of the *MYC* locus (D) was observed, thus supporting *BCL6* as the most likely translocation partner of *IGH* in those tumors. The three signals detected by the *IGH* probe (green) of the *IGH*/*BCL2* probe set (C) are consistent with the *IGH* rearrangement observed using the break-apart probe. Within each experiment, at least 100 intact, nonoverlapping nuclei were scored for each probe, and representative results are depicted.

the ones derived from the entire coding genome or from other cancer gene-containing gene sets, such as kinases and transcription factors, altogether, somatic mutations in DNA repair genes affected approximately half of DLBCL cases analyzed. Furthermore, we have provided evidence that mutations in subsets of these genes, especially those belonging to the MMR and NHEJ pathways, indeed associated with different forms of genetic instability in the tumors.

Mutations in MMR genes were associated with an MSI phenotype in DLBCL, which is a characteristic form of genetic instability in MMR-deficient tumors (Jiricny, 2006). Furthermore, these tumors displayed an overall increase of somatic mutations throughout the coding genome and of small insertions and deletions in particular. In the context of B cell development, the MMR system is involved in the mutagenic processing of mismatches introduced by AID during SHM and CSR (Martin et al., 2003; Bardwell et al., 2004; Martomo et al., 2004; Di Noia and Neuberger, 2007). Outside the *IG* locus, the MMR proteins assure a high-fidelity postreplicative DNA repair but also protect the genome from the off-target activity of AID (Jiricny, 2006; Liu et al., 2008). The distinct roles of the MMR system in SHM and postreplicative repair are largely unknown but might be explained, at least partially, by the involvement of different polymerases during these processes (Di Noia and Neuberger, 2007). The MSI and mutator phenotypes of the MMR-mutated DLBCL cases presented here might derive from both postreplicative MMR deficiency and inefficient repair of off-target AID activity. Higher mutation loads in MMR-mutated tumors, as well as the specific increase of small insertions and deletions, have been previously described in human cancers such as colorectal (Cancer Genome Atlas Network, 2012) or gastric (Nagarajan et al., 2012) cancers but never in DLBCL. The increased observation of small insertions and deletions is particularly relevant as a result of the deleterious consequences of these mutations. Similarly, to the best of our knowledge, MSI phenotypes were not previously associated with somatic mutations in MMR in DLBCL. Interestingly, we discovered that C:G→A:T transversions were relatively more frequent in MMR-mutated DLBCLs with MSI. The relative increase in C:G→A:T transversions, in the context of MMR deficiency, has been previously reported for *mutL* and *mutS* *E. coli* mutants (Zhao and Winkler, 2000) and in *Msh2* heterozygous mutant mice exposed to PhIP, a food-borne carcinogen which induces a C:G→A:T mutation signature (Zhang et al., 2001). The fact that MMR defects correlated with characteristic mutagenic signatures supports a role for MMR in DLBCL lymphomagenesis.

In the current work, translocations affecting the *IGH* locus were encountered in DLBCL cases that harbored somatic mutations in NHEJ genes. NHEJ defects, in *Trp53*-deficient mice, strongly predispose to the development of pro-B cell lymphomas that harbor interchromosomal translocations involving the *IgH* locus (Ferguson and Alt, 2001; Rooney et al., 2004). The role of C-NHEJ as a suppressor of interchromosomal translocations might relate to the fact

that it preferentially joins DSBs arising within the same chromosome (Boboila et al., 2010). The joining of interchromosomal breaks is thought to be instead mediated by a microhomology-based alternative end-joining (A-EJ) pathway (Boboila et al., 2012). Somatic defects in C-NHEJ components, which would provide a genetic basis for the occurrence of *IG*-associated translocations, have been rarely reported in human lymphomas. We previously described a somatic heterozygous mutation in the *NHEJ1* gene, encoding the Cernunnos/XLF protein in a DLBCL tumor (Du et al., 2012). This tumor sample presented concurrent translocations of *BCL2* and *MYC* with both *IGH* alleles, suggesting a relation between an NHEJ defect and the occurrence of *IGH*-associated genomic instability. We provide here further support for an association between somatic alterations in NHEJ genes (*DCLRE1C* and *XRCC5*) and the occurrence of genomic instability events involving the *IGH* locus.

The NHEJ pathway, in response to DNA DSB, is highly dependent on general damage response mechanisms. DDR proteins such as ATM, TP53BP1, and NBN are known to cooperate with NHEJ factors during V(D)J recombination and CSR (Kotnis et al., 2009; Lieber, 2010). In the current work, we discovered several novel mutations and rare variants in DDR genes, including *CHEK2* and *PARP1*, which are novel mutation targets in DLBCL. Certain germline variants in *CHEK2* are known to confer susceptibility to cancer, particularly of breast and prostate origin, whereas somatic alterations are rarely observed in this gene (Antoni et al., 2007). *CHEK2* is a downstream effector of ATM and responsible for conveying DDR signals to substrates involved in cell cycle progression and apoptosis such as the TP53 tumor suppressor and the CDC25 phosphatase family (Ahn et al., 2004). Therefore, *CHEK2* defects might foster the accumulation of DNA damage during lymphomagenesis by impairing cell cycle checkpoints and apoptotic responses. Interestingly, *CHEK2* was previously shown to be activated by the DNA-PKcs and to be associated with the XRCC5/Ku80 and XRCC6/Ku70 proteins (Li and Stern, 2005). Its potential role during the damage response to DNA DSBs during V(D)J recombination and CSR requires further investigation.

We found the *PARP1* gene to be targeted by different mechanisms that included mutations and allelic imbalance. PARP1 is involved the initial phases of DDR by accumulating at sites of damage, inducing chromatin remodeling and attracting DNA repair factors through the synthesis of poly(ADP-ribose) groups (Krishnakumar and Kraus, 2010). It has also been suggested to be a component of the A-EJ pathway (Audebert et al., 2004), particularly during CSR (Robert et al., 2009). High PARP1 activity, in response to severe DNA damage, was shown to induce cell death, most likely through energy depletion (Krishnakumar and Kraus, 2010). The targeting of PARP1 by some DLBCL tumors might, thus, constitute a survival adaptation to high degrees of genomic instability in B cells. The essential role of PARP1 in HR-deficient (BRCA2 deficient) cells has, however, stimulated the development of PARP-targeted therapies (Bryant et al., 2005).

Several clinical trials, which include B cell lymphoma patients, are currently ongoing to test the efficacy of PARP inhibitors as therapeutic agents (Kummar et al., 2011). The finding that a subset of DLBCL samples presents potentially loss-of-function mutations in *PARP1*, and/or lower expression levels of this gene supports the stratification of patients before the application of PARP-targeted therapies.

In addition to somatic mutations, several novel germline mutations were encountered in different DNA repair pathways, and we propose that the affected genes, especially the MMR genes, are good candidates for cancer susceptibility genes in mature B cell lymphomas. Heterozygous carriers of pathogenic mutations in certain MMR genes are strongly predisposed to colorectal but also to a variety of extracolonic neoplasias (Peltomäki, 2005). Although lymphomas are currently not considered to be part of the Lynch syndrome cancer spectrum, patients with biallelic mutations in certain MMR genes are predisposed to lymphomas and were sometimes shown to carry CSR defects (Wimmer and Etzler, 2008). Additionally, murine models carrying germline inactivation of MMR genes are strongly predisposed to the development of both B and T cell lymphomas (de Wind et al., 1995; Edelmann et al., 1997; Qin et al., 1999). Screening of germline variations in a larger cohort of patients and functional characterization of these variants will be required to support our hypothesis.

Deregulation of proto-oncogene expression through mutation or translocation of the *BCL6*, *BCL2*, or *MYC* genes, constitutional activation of the NF- κ B pathway through genetic lesions in *TNFAIP3*, *CARD11*, *CD79A/B*, and *MYD88* genes, and epigenetic reprogramming, triggered by mutations in genes such as *MLL2*, *EZH2*, *MEF2B*, and *CREBBP*, account for some of the most frequent events in DLBCL (Schneider et al., 2011). These alterations constitute cancer driver events, as they provide tumor cells with constitutive survival and proliferative signals, escape from apoptosis, and gene expression plasticity. Theoretically, as they all result from alterations at the DNA level, they might be a product of inefficient DNA repair mechanisms in combination with an extremely high load of DNA damage in germinal center B cells, derived from high proliferation rates and from SHM and CSR processes. Previous next-generation sequencing studies on DLBCL have not focused on DNA repair genes, possibly because of the rare observation of recurrent mutations in genes other than *TP53* (Morin et al., 2011; Pasqualucci et al., 2011; Lohr et al., 2012; Zhang et al., 2013). We report here alterations in a variety of DNA repair genes in DLBCL tumors and provide evidence for a driver effect of alterations in the MMR and NHEJ pathways. Furthermore, most of the somatic mutations found derived from dominant tumor clones, suggesting that defects in DNA repair mechanisms are important and may constitute early driver events in lymphomagenesis.

As tumor cells acquire deficiencies in DNA repair pathways they may compensate such defects by relying on other pathways that display, at least partially, functional redundancy. The targeting of compensatory pathways by therapeutic intervention, according to a principle of synthetic lethality, is

one of the major cancer drug research focuses and requires the comprehensive characterization of DNA repair defects in tumors (Curtin, 2012). Our work has provided a basis for developing similar therapeutic strategies for DLBCLs. The impact of mutations in DNA repair genes on response to treatment and patient prognosis should be further assessed in larger cohorts of patients.

MATERIALS AND METHODS

Patient materials. DNA was extracted from 29 mature B cell lymphoma frozen biopsies and the respective paired blood samples, collected at Sun Yat-Sen University Cancer Center (Guangzhou, China), to undergo targeted next-generation sequencing of DNA repair genes. This collection included 22 DLBCL, of which 19 were primary and 3 relapsed (derived from different patients), 5 primary FL, and 2 primary BL, all derived from Chinese patients. None of the patients underwent chemotherapy before biopsy collection, except for the relapsed cases where patients had been treated with R-CHOP. The median age of the DLBCL patients at diagnosis was 65 yr, and the cohort comprised 13 females and 9 males. Seven DLBCL samples were classified as germinal center B cell-like (GCB), whereas 15 DLBCL samples were classified as non-GCB, according to the Hans algorithm (Hans et al., 2004). The tumor content of the DLBCL tumors was estimated to be at least 80% in all cases by histological evaluation. DNA was also extracted from frozen biopsies from an additional 250 DLBCL, collected in biobanks at Uppsala's University Hospital (171 DLBCL from Swedish patients) and Sun Yat-Sen University Cancer Center (79 DLBCL from Chinese patients) for cohort expansion studies. 10 additional paired blood samples were available for the extended cohort of Chinese DLBCL. Furthermore, DNA isolated from the peripheral blood from 67 Chinese DLBCL patients and 429 Chinese and 1,030 Swedish blood donors was used for assessing population frequencies of a selected set of genetic variants. DNA processed by Selector technology and SOLiD sequencing was isolated by the DNeasy Tissue and Blood kit (QIAGEN), according to the manufacturer's instructions. DNA from subsets of samples composing the expanded cohort was extracted by different methodologies. RNA was extracted from 61 DLBCL cases and 8 normal lymph nodes with Trizol reagent (Life Technologies), according to the manufacturer's instructions. The institutional review boards at the Sun Yat-Sen University Cancer Center and Uppsala University approved the study.

Target enrichment and high-throughput sequencing of DNA-repair genes. 58 genomic DNAs were enriched for the coding regions of 73 DNA repair genes, as well as *TNFAIP3*, a known DLBCL tumor suppressor (Table S1). The Selector method (an earlier version of the current HaloPlex from Agilent) was applied in the enrichment step (Johansson et al., 2011). It makes use of biotin-labeled Selector probes designed to contain two sequences complementary to both ends of a DNA target. Upon hybridization, probe binding induces circularization of target sequences that can be amplified selectively. 800 ng of genomic DNA was fragmented through a combination of eight different restriction enzymatic reactions (100 ng of DNA each). After probe hybridization and circularization of the target sequences, this complex was separated from nontarget DNA by making use of streptavidin magnetic beads. After ligation of the circularized DNA molecules, specific amplification of the target sequences was achieved by a multiple displacement amplification reaction performed by a Phi29 DNA polymerase (TempliPhi; GE Healthcare).

To amplify the 1,158 exons that constituted the ROI, two different Selector designs were attempted. Target nucleotides were covered by more than one selector probe to minimize the risk of hybridization failure provoked by nucleotide variations. In design 1, each targeted base was covered on average by 2.2 selector amplicons, whereas design 2 produced shorter fragments and allowed targeted bases to be covered on average by 3.3 selector amplicons (Table S2). Noncovered sequences in the ROI were mostly composed of repeat regions as identified by the RepeatMasker database (<http://www.repeatmasker.org>). In design 2, repetitive sequences were further avoided to improve sequencing efficiency at the cost of a slight reduction in the coverage

of the ROI (Table S2). Five DLBCL and respective normal pairs were enriched by design 1, and the remaining samples were enriched by design 2. Success of enrichment of target sequences was monitored by quantitative PCR to measure the relative levels of several on- and off-target amplicons.

300 ng of each enrichment product was fragmented using the Covaris S2 system, after which the standard library construction kit for SOLiD v3 (Applied Biosystems), for design 1, and SOLiD v4, for design 2, were used to barcode and prepare the samples for sequencing. Barcoded samples were pooled and sequenced at the Uppsala Genome Center (Uppsala, Sweden). Samples from design 1 were sequenced in two separate runs, each making use of a quarter of a sequencing slide, whereas samples derived from design 2 were sequenced on three quarters of sequencing slide. The SOLiD libraries from 16 samples from design 2 were further sequenced in an additional run to increase coverage and depth. A HapMap sample NA19240 was subjected to the same enrichment (design 1) and sequencing procedures to evaluate the method's specificity and reproducibility. Design of selector probes and sequence library preparation were performed at Halo Genomics (Uppsala, Sweden).

Analysis of SOLiD data. SOLiD short reads were aligned with MosaikAligner version 1.1.0018; -hs 15, -mm 6, -ms 6, -mhp 100, and -act 20 (Mosaik; The MarthLab) against the human reference genome (hg18). Alignments from samples that were sequenced twice were merged using MosaikMerge. The combined alignments were then exported to SAM/BAM format and reanalyzed by in-house software to deal with artificial insertions/deletions provoked by "forced" alignments. Single nucleotide variants (SNVs) and small insertions and deletions (indels) detection were performed with SNPmania (in-house software). Somatic alterations were determined by comparing the MAF between tumor and normal at specific base positions. A minimum of 10 and 5 reads in either normal and tumor samples was required as well as the representation of the minor allele by at least 5% of the reads in the tumor sample. Additionally, variant detection against the human reference genome (hg18) was also performed by flagging base positions in which the reference allele was represented at a frequency of 0–0.8 in positions covered by at least 10 reads. To maximize mutation discovery, SOLiD short reads were also aligned to the hg18 reference using BFAST (Homer et al., 2009). SAMtools was used to extract base counts for each position (Li et al., 2009a). Allelic imbalance events were detected by comparing the allelic ratios at germline heterozygous single nucleotide variants between tumor and normal samples. The presence of allelic imbalance at a gene had to be supported by at least two heterozygous positions where the allelic ratios diverged significantly from the heterozygous state. Fischer's exact test was used for the comparison of single nucleotide positions between tumor and normal samples. Functional prediction of missense mutations was performed in silico with SIFT and PolyPhen-2 (Kumar et al., 2009; Adzhubei et al., 2010). Visual inspection of reads at relevant base positions was performed with the Integrative Genomics Viewer (IGV) software (Robinson et al., 2011). SNP concordance analysis for the HapMap sample NA19240 was performed with the Genome Analysis Toolkit (DePristo et al., 2011).

Sanger sequencing. Validation of variants discovered by next-generation sequencing, and sequencing of the *CHEK2*, *MSH2*, and *PARP1* genes in the cohort expansion studies, were performed by Sanger sequencing. Specific primers were designed with Primer 3 (<http://fokker.wi.mit.edu/primer3/>). PCR reactions were run in a total volume of 25 μ l containing 1 U GoTaq polymerase (Promega), 0.4 μ M of forward and reverse primers, 1.5 mM $MgCl_2$, 200 μ M dNTPs, and 10 ng DNA. PCR products were purified and sequenced at MacroGen Inc. or Eurofins MWG Operon.

454 amplicon sequencing. Screening of *DCLRE1C*, *DDB1*, *MLH3*, *XRCC4*, and *XRCC5* in an expanded cohort was performed by means of amplicon deep sequencing. Microfluidic PCR reactions were run in a 48 \times 48 Access array system (Fluidigm) with FastStart High Fidelity PCR system (Roche) according to the manufacturer's instructions. Specific primers containing universal tails (CS1 and CS2) were designed to produce amplicons between 300 and 350 bp at the coding regions of the genes of interest.

Molecularly barcoded 454 FLX Titanium (Roche) sequencing primers (Lib-A emPCR kit design), containing sequences complementary to the CS1 and CS2 primer tails, were added to the Access array together with 50 ng of each DNA sample (48 in total) so that all amplicons derived from the same DNA sample shared the same barcode. Indexed libraries were recovered for each sample in a single well of the Access array and analyzed with Agilent's 2100 Bioanalyzer (Agilent Technologies). Thereafter, samples were pooled and purified by gel extraction. After emulsion, PCR (Lib-A) samples were sequenced by the GS FLX+ System in a full PicoTiter Plate at MacroGen Inc. Analysis was performed with the GS Amplicon Variant Analyzer and potential variants validated by Sanger sequencing.

Population frequencies of single nucleotide variants. A proportion of the germline variants discovered by next-generation sequencing, together with all of the *CHEK2* variants, were genotyped in Swedish and Chinese control samples using matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF; Sequenom Inc.) as previously described (Jurinke et al., 2002). PCR assays and associated extension reactions were designed using the MassArray assay design 3.1 software (Sequenom Inc.). Primers were acquired from Metabion GmbH (Planegg-Martinsried). The extension products were analyzed by the MassArray Compact MALDI-TOF mass spectrometer (Sequenom Inc.). The MassARRAY Typer 4.0 software (Sequenom Inc.) was used to analyze the resulting mass spectra for peak identification. Two independent scorers confirmed all genotypes. For the c.1100delC, p.H371Y and p.D438Y *CHEK2* variants control typing was performed with Sanger sequencing because of the inability to produce a specific assay for MALDI-TOF. As the population-specific MAFs determined by typing were highly concordant with the ones derived from the 1000 Genomes (<http://www.1000genomes.org/>) sequencing data, the latter were subsequently used for determining the MAFs of all germline variants discovered by next generation sequencing. MAFs were extracted from 1000 Genomes with the aid of VCFtools (Danecek et al., 2011).

Allelic imbalance and MSI analysis. Polymorphic sequence-tagged site (STS) markers were used to determine the zygosity status of the *EXO1*, *MDC1*, *PARP1*, *RPA1*, and *TP53BP1* loci. STS markers and respective primer sequences were retrieved from the UCSC genome browser (<http://genome.ucsc.edu/>) or the Ensembl genome browser (<http://www.ensembl.org/>) and, when necessary, primers were redesigned with Primer 3. The mononucleotide microsatellite markers BAT-25 and BAT-26 and the dinucleotide markers D2S123 and D10S197 were used for the detection of MSI. Forward primers were fluorescently labeled with 6-carboxyfluorescein (6-FAM) at their 5'-end and purified by HPLC. PCR reactions were run in a total of 15 μ l, containing 0.5 U GoTaqDNA polymerase (Promega), 0.66 μ M of each primer, 330 μ M dNTPs, 10 ng DNA, and 1.5 mM $MgCl_2$. PCR products were diluted appropriately and mixed in a formamide solution containing GeneScan 500 LIZ Size Standard (Applied Biosystems). Thereafter, samples were loaded in a 48-capillary 3730 DNA Analyzer and analyzed with Peak Scanner software version 1.0 (Applied Biosystems). For the *MSH2*-mutated samples, MSI testing was also performed using a commercial kit (Microsatellite Instability Multiplex System kit; Promega) according to the manufacturer's instructions. In brief, genomic DNA prepared from tumor tissue was amplified by PCR in a multiplex PCR, amplifying the selected microsatellite markers. Microsatellite analysis was performed using an ABI3500xl (Applied Biosystems) with the GeneMapper 4.1 software (Applied Biosystems).

Gene expression analysis. The relative expression of the *EXO1*, *MDC1*, *PARP1*, *RPA1*, and *TP53BP1* genes was assessed by quantitative real time PCR (qPCR). Furthermore, the expression levels of the housekeeping genes *ACTB*, *GAPDH*, and *HRPT1* were determined to perform inter-sample normalization. Intron-spanning assays were designed with Primer 3. RNAs were processed with the first-strand cDNA synthesis kit (GE Healthcare) and 10- μ l reactions were prepared with the KAPA SYBR FAST ABI Prism qPCR reagent (Kapa Biosystems). Samples and standard curves were run in duplicate in a 7500 Fast Real-Time PCR System (Applied Biosystems). Analysis was performed with the 7500 System SDS Software, version 1.4 (Applied Biosystems).

Exome sequencing. Exome sequencing was performed in 19 DLBCL tumors, including 17 tumors that underwent targeted sequencing of DNA repair genes and 2 additional tumors that displayed MSI. In brief, genomic DNA was processed into fragments between 150 and 200 bp and Illumina sequencing adapters were ligated to both ends of the resulting fragments. The adapter-ligated templates were purified by AgencourtAMPure SPRI beads (Beckman Coulter) and fragments with insert size of ~200 bp were excised. Extracted DNA was amplified by ligation-mediated polymerase chain reaction (LM-PCR), purified, and hybridized to the SureSelectBiotinylated RNA Library (BAITS; Agilent Technologies) for enrichment. Hybridized fragments were bound to streptavidin beads, whereas nonhybridized fragments were washed out after 24 h of incubation. Captured LM-PCR products were then subjected to the 2100 Bioanalyzer (Agilent) to estimate the magnitude of enrichment. Each captured library was loaded on a HiSeq2000 platform (Illumina), and high-throughput sequencing was performed for each captured library independently to ensure that each sample met the desired mean fold coverage (>50×). Raw image files were processed by base calling software (1.7; Illumina) for base calling with default parameters and the sequences of each individual were generated as 90-bp paired-end reads. During bioinformatic analysis, the adapter sequences in the raw data were removed, and low-quality reads were discarded. Sequences were aligned to the genome with Burrows-Wheeler Aligner (BWA) from which BAM files were obtained (Li and Durbin, 2009). Somatic indels were detected with GATK (Li et al., 2009b), whereas SNVs were detected using Varscan (Koboldt et al., 2009). All variants underwent visual inspection. Sequencing and data analysis was performed at BGI (Shenzhen, China).

Detection of chromosomal translocations by FISH. Vysis LSI IGH/BCL2 dual-color, dual-fusion translocation probe, Vysis LSI BCL6 dual-color, break-apart rearrangement probe, Vysis LSI MYC dual-color, break-apart rearrangement probe, and Vysis LSI IGH break-apart rearrangement probe were purchased from Abbott Molecular Inc. Sample processing and hybridization were performed according to the manufacturer's instruction and, as described previously, (Ribeiro et al., 2006). Slides were counterstained with DAPI (Vector Laboratories), and fluorescent images were sequentially captured with a Cohu 4900 charge-coupled device camera coupled to an Axioplan fluorescence microscope (Carl Zeiss).

Statistical analysis. Statistical tests and graph construction was performed with Prism (version 4.00; GraphPad Software).

Online supplemental material. Table S1 shows the genes sequenced by Selector and SOLiD. Table S2 summarizes the Selector capture and SOLiD sequencing performances. Table S3 shows all somatic and germline (novel and rare) mutations identified by Selector/SOLiD sequencing. Fig. S1 summarizes the SNV detection process from the Selector/SOLiD sequencing data. Online supplemental material is available at <http://www.jem.org/cgi/content/full/jem.20122842/DC1>.

We thank Dr. K. Duvefelt for help in genotyping, R. Vossen for assistance with the Fluidigm technology, and Prof. L. Hammarström for critical reading of the manuscript.

This work was supported by the Swedish Cancer Society, the Swedish Research Council, the European Research Council (242551-ImmunoSwitch), and the KIDS program at the Karolinska Institutet.

The authors have no conflicting financial interests.

Submitted: 24 December 2012

Accepted: 24 July 2013

REFERENCES

Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, and S.R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods*. 7:248–249. <http://dx.doi.org/10.1038/nmeth0410-248>

- Ahn, J., M. Urist, and C. Prives. 2004. The Chk2 protein kinase. *DNA Repair (Amst.)*. 3:1039–1047. <http://dx.doi.org/10.1016/j.dnarep.2004.03.033>
- Antoni, L., N. Sodha, I. Collins, and M.D. Garrett. 2007. CHK2 kinase: cancer susceptibility and cancer therapy – two sides of the same coin? *Nat. Rev. Cancer*. 7:925–936. <http://dx.doi.org/10.1038/nrc2251>
- Audebert, M., B. Salles, and P. Calsou. 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J. Biol. Chem.* 279: 55117–55126. <http://dx.doi.org/10.1074/jbc.M404524200>
- Bardwell, P.D., C.J. Woo, K. Wei, Z. Li, A. Martin, S.Z. Sack, T. Parris, W. Edelmann, and M.D. Scharff. 2004. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat. Immunol.* 5:224–229. <http://dx.doi.org/10.1038/ni1031>
- Bell, D.W., S.H. Kim, A.K. Godwin, T.A. Schiripo, P.L. Harris, S.M. Haserlat, D.C. Wahrer, C.A. Haiman, M.B. Daly, K.B. Niendorf, et al. 2007. Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. *Int. J. Cancer*. 121:2661–2667. <http://dx.doi.org/10.1002/ijc.23026>
- Boboila, C., M. Jankovic, C.T. Yan, J.H. Wang, D.R. Wesemann, T. Zhang, A. Fazeli, L. Feldman, A. Nussenzweig, M. Nussenzweig, and F.W. Alt. 2010. Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. *Proc. Natl. Acad. Sci. USA*. 107:3034–3039. <http://dx.doi.org/10.1073/pnas.0915067107>
- Boboila, C., F.W. Alt, and B. Schwer. 2012. Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv. Immunol.* 116:1–49. <http://dx.doi.org/10.1016/B978-0-12-394300-2.00001-6>
- Bryant, H.E., N. Schultz, H.D. Thomas, K.M. Parker, D. Flower, E. Lopez, S. Kyle, M. Meuth, N.J. Curtin, and T. Hellday. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*. 434:913–917. <http://dx.doi.org/10.1038/nature03443>
- Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 487:330–337. <http://dx.doi.org/10.1038/nature11252>
- Compagno, M., W.K. Lim, A. Grunn, S.V. Nandula, M. Brahmachary, Q. Shen, F. Bertoni, M. Ponzoni, M. Scandurra, A. Califano, et al. 2009. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*. 459:717–721. <http://dx.doi.org/10.1038/nature07968>
- Curtin, N.J. 2012. DNA repair dysregulation from cancer driver to therapeutic target. *Nat. Rev. Cancer*. 12:801–817. <http://dx.doi.org/10.1038/nrc3399>
- Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, et al. 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156–2158. <http://dx.doi.org/10.1093/bioinformatics/btr330>
- de Miranda, N.F., A. Björkman, and Q. Pan-Hammarström. 2011. DNA repair: the link between primary immunodeficiency and cancer. *Ann. N.Y. Acad. Sci.* 1246:50–63. <http://dx.doi.org/10.1111/j.1749-6632.2011.06322.x>
- de Wind, N., M. Dekker, A. Berns, M. Radman, and H. te Riele. 1995. Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell*. 82:321–330. [http://dx.doi.org/10.1016/0092-8674\(95\)90319-4](http://dx.doi.org/10.1016/0092-8674(95)90319-4)
- DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, A.A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–498. <http://dx.doi.org/10.1038/ng.806>
- Di Noia, J.M., and M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* 76:1–22. <http://dx.doi.org/10.1146/annurev.biochem.76.061705.090740>
- Du, L., R. Peng, A. Björkman, N. Filipe de Miranda, C. Rosner, A. Kotnis, M. Berglund, C. Liu, R. Rosenquist, G. Enblad, et al. 2012. Cernunnos influences human immunoglobulin class switch recombination and may be associated with B cell lymphomagenesis. *J. Exp. Med.* 209:291–305. <http://dx.doi.org/10.1084/jem.20110325>
- Edelmann, W., K. Yang, A. Umar, J. Heyer, K. Lau, K. Fan, W. Liedtke, P.E. Cohen, M.F. Kane, J.R. Lipford, et al. 1997. Mutation in the

- mismatch repair gene Msh6 causes cancer susceptibility. *Cell*. 91:467–477. [http://dx.doi.org/10.1016/S0092-8674\(00\)80433-X](http://dx.doi.org/10.1016/S0092-8674(00)80433-X)
- Falck, J., N. Maitland, R.G. Syljuåsen, J. Bartek, and J. Lukas. 2001. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature*. 410:842–847. <http://dx.doi.org/10.1038/35071124>
- Ferguson, D.O., and F.W. Alt. 2001. DNA double strand break repair and chromosomal translocation: lessons from animal models. *Oncogene*. 20:5572–5579. <http://dx.doi.org/10.1038/sj.onc.1204767>
- Guidi, C.J., T.M. Veal, S.N. Jones, and A.N. Imbalzano. 2004. Transcriptional compensation for loss of an allele of the *Ini1* tumor suppressor. *J. Biol. Chem.* 279:4180–4185. <http://dx.doi.org/10.1074/jbc.M312043200>
- Hans, C.P., D.D. Weisenburger, T.C. Greiner, R.D. Gascoyne, J. Delabie, G. Ott, H.K. Müller-Hermelink, E. Campo, R.M. Braziel, E.S. Jaffe, et al. 2004. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 103:275–282. <http://dx.doi.org/10.1182/blood-2003-05-1545>
- Hicks, S., D.A. Wheeler, S.E. Plon, and M. Kimmel. 2011. Prediction of mis-sense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum. Mutat.* 32:661–668. <http://dx.doi.org/10.1002/humu.21490>
- Homer, N., B. Merriman, and S.F. Nelson. 2009. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE*. 4:e7767. <http://dx.doi.org/10.1371/journal.pone.0007767>
- Jiricny, J. 2006. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 7:335–346. <http://dx.doi.org/10.1038/nrm1907>
- Johansson, H., M. Isaksson, E.F. Sörqvist, F. Roos, J. Stenberg, T. Sjöblom, J. Botling, P. Micke, K. Edlund, S. Fredriksson, et al. 2011. Targeted resequencing of candidate genes using selector probes. *Nucleic Acids Res.* 39:e8. <http://dx.doi.org/10.1093/nar/gkq1005>
- Jung, D., C. Giallourakis, R. Mostoslavsky, and F.W. Alt. 2006. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* 24:541–570. <http://dx.doi.org/10.1146/annurev.immunol.23.021704.115830>
- Jurinke, C., D. van den Boom, C.R. Cantor, and H. Köster. 2002. Automated genotyping using the DNA MassArray technology. *Methods Mol. Biol.* 187:179–192.
- Klein, U., and R. Dalla-Favera. 2008. Germinal centres: role in B-cell physiology and malignancy. *Nat. Rev. Immunol.* 8:22–33. <http://dx.doi.org/10.1038/nri2217>
- Koboldt, D.C., K. Chen, T. Wylie, D.E. Larson, M.D. McLellan, E.R. Mardis, G.M. Weinstock, R.K. Wilson, and L. Ding. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 25:2283–2285. <http://dx.doi.org/10.1093/bioinformatics/btp373>
- Kotnis, A., L. Du, C. Liu, S.W. Popov, and Q. Pan-Hammarström. 2009. Non-homologous end joining in class switch recombination: the beginning of the end. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:653–665. <http://dx.doi.org/10.1098/rstb.2008.0196>
- Krishnakumar, R., and W.L. Kraus. 2010. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell*. 39:8–24. <http://dx.doi.org/10.1016/j.molcel.2010.06.017>
- Kumar, P., S. Henikoff, and P.C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4:1073–1081. <http://dx.doi.org/10.1038/nprot.2009.86>
- Kummar, S., A. Chen, J. Ji, Y. Zhang, J.M. Reid, M. Ames, L. Jia, M. Weil, G. Speranza, A.J. Murgo, et al. 2011. Phase I study of PARP inhibitor ABT-888 in combination with topotecan in adults with refractory solid tumors and lymphomas. *Cancer Res.* 71:5626–5634. <http://dx.doi.org/10.1158/0008-5472.CAN-11-1227>
- Küppers, R., and R. Dalla-Favera. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. 20:5580–5594. <http://dx.doi.org/10.1038/sj.onc.1204640>
- Lenz, G., I. Nagel, R. Siebert, A.V. Roschke, W. Sanger, G.W. Wright, S.S. Dave, B. Tan, H. Zhao, A. Rosenwald, et al. 2007. Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *J. Exp. Med.* 204:633–643. <http://dx.doi.org/10.1084/jem.20062041>
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin; 1000 Genome Project Data Processing Subgroup. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>
- Li, R., Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, and J. Wang. 2009b. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* 19:1124–1132. <http://dx.doi.org/10.1101/gr.088013.108>
- Li, J., and D.F. Stern. 2005. Regulation of CHK2 by DNA-dependent protein kinase. *J. Biol. Chem.* 280:12041–12050. <http://dx.doi.org/10.1074/jbc.M412445200>
- Lieber, M.R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79:181–211. <http://dx.doi.org/10.1146/annurev.biochem.052308.093131>
- Liu, H.X., X.L. Zhou, T. Liu, B. Werelius, G. Lindmark, N. Dahl, and A. Lindblom. 2003. The role of hMLH3 in familial colorectal cancer. *Cancer Res.* 63:1894–1899.
- Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature*. 451:841–845. <http://dx.doi.org/10.1038/nature06547>
- Lohr, J.G., P. Stojanov, M.S. Lawrence, D. Auclair, B. Chapuy, C. Sougnez, P. Cruz-Gordillo, B. Knoechel, Y.W. Asmann, S.L. Slager, et al. 2012. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. USA*. 109:3879–3884. <http://dx.doi.org/10.1073/pnas.1121343109>
- Martin, A., Z. Li, D.P. Lin, P.D. Bardwell, M.D. Iglesias-Ussel, W. Edelmann, and M.D. Scharff. 2003. Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. *J. Exp. Med.* 198:1171–1178. <http://dx.doi.org/10.1084/jem.20030880>
- Martomo, S.A., W.W. Yang, and P.J. Gearhart. 2004. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J. Exp. Med.* 200:61–68. <http://dx.doi.org/10.1084/jem.20040691>
- Morin, R.D., M. Mendez-Lago, A.J. Mungall, R. Goya, K.L. Mungall, R.D. Corbett, N.A. Johnson, T.M. Severson, R. Chiu, M. Field, et al. 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 476:298–303. <http://dx.doi.org/10.1038/nature10351>
- Nagarajan, N., D. Bertrand, A.M. Hillmer, Z.J. Zang, F. Yao, P.E. Jacques, A.S. Teo, I. Cutcutache, Z. Zhang, W.H. Lee, et al. 2012. Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome Biol.* 13:R115. <http://dx.doi.org/10.1186/gb-2012-13-12-r115>
- Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R.S. Chaganti, R. Küppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 412:341–346. <http://dx.doi.org/10.1038/35085588>
- Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H.C. Morse III, M.C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. *Nat. Genet.* 40:108–112. <http://dx.doi.org/10.1038/ng.2007.35>
- Pasqualucci, L., V. Trifonov, G. Fabbri, J. Ma, D. Rossi, A. Chiarenza, V.A. Wells, A. Grunn, M. Messina, O. Elliot, et al. 2011. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* 43:830–837. <http://dx.doi.org/10.1038/ng.892>
- Peltomäki, P. 2005. Lynch syndrome genes. *Fam. Cancer*. 4:227–232. <http://dx.doi.org/10.1007/s10689-004-7993-0>
- Qin, X., L. Liu, and S.L. Gerson. 1999. Mice defective in the DNA mismatch gene PMS2 are hypersensitive to MNU induced thymic lymphoma and are partially protected by transgenic expression of human MGMT. *Oncogene*. 18:4394–4400. <http://dx.doi.org/10.1038/sj.onc.1202798>
- Ramiro, A.R., M. Jankovic, T. Eisenreich, S. Difilippantonio, S. Chen-Kiang, M. Muramatsu, T. Honjo, A. Nussenzweig, and M.C. Nussenzweig. 2004. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*. 118:431–438. <http://dx.doi.org/10.1016/j.cell.2004.08.006>
- Ribeiro, F.R., C. Jerónimo, R. Henrique, D. Fonseca, J. Oliveira, R.A. Lothe, and M.R. Teixeira. 2006. 8q gain is an independent predictor of poor survival

- in diagnostic needle biopsies from prostate cancer suspects. *Clin. Cancer Res.* 12:3961–3970. <http://dx.doi.org/10.1158/1078-0432.CCR-05-1977>
- Robert, I., F. Dantzer, and B. Reina-San-Martin. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J. Exp. Med.* 206:1047–1056. <http://dx.doi.org/10.1084/jem.20082468>
- Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26. <http://dx.doi.org/10.1038/nbt.1754>
- Rooney, S., J. Sekiguchi, S. Whitlow, M. Eckersdorff, J.P. Manis, C. Lee, D.O. Ferguson, and F.W. Alt. 2004. Artemis and p53 cooperate to suppress oncogenic N-myc amplification in progenitor B cells. *Proc. Natl. Acad. Sci. USA.* 101:2410–2415. <http://dx.doi.org/10.1073/pnas.0308757101>
- Schneider, C., L. Pasqualucci, and R. Dalla-Favera. 2011. Molecular pathogenesis of diffuse large B-cell lymphoma. *Semin. Diagn. Pathol.* 28:167–177. <http://dx.doi.org/10.1053/j.semdp.2011.04.001>
- Sodha, N., T.S. Mantoni, S.V. Tavtigian, R. Eeles, and M.D. Garrett. 2006. Rare germ line CHEK2 variants identified in breast cancer families encode proteins that show impaired activation. *Cancer Res.* 66:8966–8970. <http://dx.doi.org/10.1158/0008-5472.CAN-06-1990>
- Stavnezer, J., A. Björkman, L. Du, A. Cagigi, and Q. Pan-Hammarström. 2010. Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Adv. Immunol.* 108:45–109. <http://dx.doi.org/10.1016/B978-0-12-380995-7.00003-3>
- Wimmer, K., and J. Eitzler. 2008. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? *Hum. Genet.* 124:105–122. <http://dx.doi.org/10.1007/s00439-008-0542-4>
- Zhang, J., V. Grubor, C.L. Love, A. Banerjee, K.L. Richards, P.A. Mieczkowski, C. Dunphy, W. Choi, W.Y. Au, G. Srivastava, et al. 2013. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. USA.* 110:1398–1403. <http://dx.doi.org/10.1073/pnas.1205299110>
- Zhang, S., R. Lloyd, G. Bowden, B.W. Glickman, and J.G. de Boer. 2001. Msh2 DNA mismatch repair gene deficiency and the food-borne mutagen 2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine (PhIP) synergistically affect mutagenesis in mouse colon. *Oncogene.* 20:6066–6072.
- Zhao, J. and M.E. Winkler. 2000. Reduction of GC → TA transversion mutation by overexpression of MutS in *Escherichia coli* K-12. *J. Bacteriol.* 182:5025–5028. <http://dx.doi.org/10.1128/JB.182.17.5025-5028.2000>